Computational and Structural Biotechnology Journal 21 (2023) 4974–4987

Contents lists available at ScienceDirect

Computational and Structural Biotechnology Journal

journal homepage: www.elsevier.com/locate/csbj

Database article

PlantCHRs: A comprehensive database of plant chromatin remodeling factors

Hengyu Yan [a,*,1], Fangyuan Liu [a,1], Guowei Zhang [a,1], Shuai Liu [a], Weiwei Ma [a], Ting Yang [a], Yubin Li [a], Jiaotong Yang [b], Hailong Cui [c,*]

[a] College of Agronomy, Qingdao Agricultural University, China
[b] Resource Institute for Chinese and Ethnic Materia Medica, Guizhou University of Traditional Chinese Medicine, Guiyang, China
[c] College of Economics and Management (Cooperative College), Qingdao Agricultural University, China

ARTICLE INFO

Keywords:
Snf2
Chromatin remodeling factors
Database
Plant
Evolution

ABSTRACT

The Snf2 protein family is a group of ATP-dependent chromatin remodeling factors (CHRs) that play an essential role in gene expression regulation. In plants, Snf2 is involved in growth, development, as well as stress resistance. However, only a very limited number of experimentally validated Snf2 have been identified and reported, while the majority remaining undiscovered in most species . In this study, we predicted 3135 Snf2 proteins and 8398 chromatin remodeling complex (CRC) subunits in diverse plant species, and constructed the Plant Chromatin Remodeling Factors Database (PlantCHRs, http://www.functionalgenomics.cn/PlantCHRs/), which provide a comprehensive resource for researchers to access information about plant CHRs. We also developed an online tool capable of predicting CHRs and CRC subunits. Moreover, we investigated the distribution of Snf2 proteins in different species and observed a significant increase in the number of Snf2 proteins and the diversity of the Snf2 subfamily during the evolution, highlighting their evolutionary importance. By analyzing the expression patterns of the Snf2 genes in different tissues of maize and *Arabidopsis*, we found that the Snf2 proteins may show some conservation across different species in regulating plant growth and development. Over the all, we established a comprehensive database for plant CHRs, which will facilitate the researches on plant chromatin remodeling.

1. Introduction

Chromatin, a complex of DNA and protein in eukaryotic cells, consists of nucleosomes as its fundamental unit. The dense arrangement of nucleosomes packages chromosomal DNA into a compact structure, effectively storing genetic information in chromatin while obstructing the access of transcription machinery. Consequently, eukaryotes evolved Snf2 proteins as CHRs that involved in biological processes, such as gene transcription, DNA repair, DNA replication by altering the position of nucleosomes. DNA transcription is primarily carried out by RNA polymerase, which first binds to the DNA promoter with the assistance of transcription factors to form a transcription initiation complex. The complex consumes nucleotides and continuously extends in the 3′ direction. In eukaryotes, DNA is wrapped around histones to form nucleosomes, a structure that protects the DNA. Therefore, prior to DNA expression in eukaryotes, CHRs are recruited to the vicinity of the DNA that needs to be expressed. These factors move the DNA out of the histones through ATP hydrolysis, and may even remove the histones themselves [1,2]. Once the transcription is complete, the CHRs could add the removed nucleosomes back onto the nucleosome string. It has been discovered that CHRs also have the ability to replace H2A and organize the nucleosomes. SWR1, for example, is capable of replacing H2A with H2A.Z, which then promotes the transition from euchromatin to heterochromatin [3]. Additionally, during the process of DNA replication, CHRs are required to orderly arrange the positioning of nucleosomes, allowing chromatin to fold into the normal 30 nm chromatin fiber with the help of other proteins. Based on a multiple sequence alignment of approximately 1300 Snf2 proteins, Flaus and colleagues categorized Snf2 proteins into 24 subfamilies: ALC1, ISWI, CHD7, Chd1, Mi-2, Lsh1, Snf2, EP400, Swr1, Ino80, Etl1, ERCC6, SSO 1653, Mot1, ATRX, Arip4, JBP2, DRD1, Rad54, Lodestar, Rad5/16, Ris1, SHPRH, SMARCAL1 [4,5]. Snf2 proteins comprising two conserved domains: SNF2-N and helicase-C, are extensively present in eukaryotes. Protein domains such as HAND, SANT, PHD, CHROMO, and HSA also were

* Corresponding authors.
E-mail addresses: hengyuyan@qau.edu.cn (H. Yan), cuihailong1977@163.com (H. Cui).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.csbj.2023.10.005
Received 8 June 2023; Received in revised form 3 October 2023; Accepted 4 October 2023
Available online 5 October 2023

Table 1
Statistics of the number of different SNF2 proteins in the collected data.

Subfamily	Non-plant organisms	*Arabidopsis*	Tomato	Rice	Total
ALC1	4	1	1	1	7
Arip4	3				3
ATRX	5	1	1	1	8
Chd1	5	1	2	1	9
CHD7	15				15
DRD1		6	11	7	24
EP400	2				2
ERCC6	8	3	3	4	18
Etl1	6	1	1	1	9
Ino80	28	1	1	1	31
ISWI	7	2	2	2	13
JBP2	6				6
Lodestar	3				3
Lsh1	5	1	1	1	8
Mi-2	9	3	2	3	17
Mot1	5	1	1	1	8
Rad5/16	17	5	5	4	31
Rad54	21	1	1	1	24
Ris1	1	5	5	3	14
SHPRH	5	2	1	2	10
SMARCAL1	6	2	2	2	12
Snf2	8	4	3	3	18
SSO1653	1				1
Swr1	16	1	1	1	19

identified in Snf2 proteins. These domains may contribute to SNF2 protein functional specialization [6].

Snf2 proteins play crucial roles in plant growth, development, and stress response. The *Arabidopsis* genome encodes 41 Snf2 genes, which can be classified into various subfamilies. The Snf2 subfamily consists of four members: *BRM*, *SYD*, *MINU1*, and *MINU2*. Mutants of *BRM* and *SYD* both exhibit dwarfism, leaf curling, and early flowering, with the *brm* mutant also showing enhanced drought tolerance [7-9]. The *MINU1* and *MINU2* genes are involved in embryonic development, with the *minu1-minu2* double mutant exhibiting impaired embryonic development [10]. Overexpression of *MINU1* leads to growth inhibition under stress conditions, while *MINU1*-deficient mutants show tolerance to salt, drought, and high temperatures [11]. The ISWI subfamily includes two members, *CHR11* and *CHR17*. The *chr11chr17* double mutants display growth and developmental defects, such as dwarfism, early flowering, curly rosette leaves, shortened primary roots, and an accelerated transition from the vegetative to the reproductive phase. Additionally, many defense response-related genes are upregulated in these double mutants in the absence of pathogen infection. The *chr11* single mutant exhibits enhanced resistance against Pst DC3000 [12,13]. The Mi-2 subfamily member *PKL* is involved in regulating various developmental processes, including embryonic development, seed germination, and temperature-mediated seedling development [14]. The DDM1 gene, belonging to the Lsh1 subfamily, regulates the resistance of *Arabidopsis thaliana* to pathogens by influencing the levels of DNA methylation in stress-responsive genes [15]. CHR19, a member of the Etl1 subfamily, plays a crucial role in regulating the SA and JA signaling pathways. Mutations in *CHR19* increase susceptibility to the necrotrophic fungal pathogen Botrytis cinerea through the JA pathway, while having no impact on the growth of the hemibiotrophic bacterial pathogen Pst DC3000 via the SA pathway [16]. RAD54 is necessary for RAD51-mediated meiotic DNA double-strand break repair [17]. Mutations in *PIE1*, a member of the Swr1 subfamily, suppress the delay in flowering mediated by FLC. PIE1 forms a complex with ARP6, contributing to the regulation of H2A.Z deposition, thus influencing the expression of specific genes [18]. EDA16, belonging to the Ris1 subfamily, functions as a negative regulator of immune response [19]. In addition to *Arabidopsis*, Snf2 proteins have been studied in other plant species. In maize, ZmCHB101, an Snf2 protein, is involved in the regulation of nitrate response. *ZmCHB101-RNAi* lines exhibit accelerated root growth and increased biomass under low nitrate conditions [20]. This finding suggests that ZmCHB101 plays a crucial role in maize adaptation to low-nitrogen environments. In rice, *RFS*, a Mi-2-like CHR, regulates multiple biological processes, including chloroplast development, leaf rolling, and reactive oxygen species scavenging (ROS) [21]. Another example in rice is OsALT1, a Ris1 subfamily chromatin remodeling ATPase. The *alt1* mutants enhance rice tolerance to alkaline stress by reducing ROS levels and alleviating oxidative stress damage [22]. A recent study suggests that OsCHR11 regulates the resistance to disease by influencing the occupancy of nucleosomes [23]. This finding suggests that Snf2 proteins play a significant role in enhancing stress tolerance in crops. Together, these studies on Snf2 proteins in different plant species underscore Snf2 roles in plant development, growth, and environmental stress response.

The development of high-throughput sequencing technologies led to genome sequencing for numerous plants, which were stored in various public platforms such as Phytozome, Ensemble Plant, and NCBI [24,25]. These resources have enabled scientists to establish specialized databases to store and analyze specific factors, such as the Plant Transcription Factor Database (PlantTFDB) [26], Leaf Senescence Database (LSD) [27], and Plant Ethylene Response Factor Database (PlantEAR) [28], effectively promoting research in these filed. Snf2 proteins, as important regulatory factors in various biological processes, are crucial for understanding the relationships between organisms and their environments, as well as for breeding stress-resistant crops. Although some databases of CHRs have been published, these databases mainly focus on the Snf2 proteins of humans and other animals, and there are no databases to classify and store plant Snf2 proteins. To facilitate the study of plant Snf2 protein, a new database called the Plant Chromatin Remodeling Factors Database (PlantCHRs) has been constructed. The database predicts CHRs and CRC subunits in different species and integrates information on protein domains, Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG), and protein-protein interactions (PPI). By providing a comprehensive platform for the analysis and comparison of Snf2 proteins across different species, PlantCHRs has the potential to greatly advance our understanding of the roles of these proteins in plant growth, development, and stress response.

2. Materials and methods

2.1. Data sources

To construct Hidden Markov Model (HMM) profiles of the Snf2 subfamily, well-annotated Snf2 proteins from different species were collected. Snf2 proteins from non-plant species were obtained from various sources, including the UniProtKB database [29], EpiFactors database [30], ChromoHub database [31], and several publications [32]. Detailed information on Snf2 proteins from non-plant species can be found in Supplementary Table 1. Snf2 proteins from *Arabidopsis thaliana*, rice, and tomato were derived from previous studies [33,34]. Detailed information on Snf2 proteins from these plant species is available in Supplementary Table 2. Snf2 proteins typically function by forming CRCs with other proteins. To predict CRCs in other plant species, CRCs of *Arabidopsis thaliana* were collected from literatures [35]. Protein sequences, transcript sequences, coding sequences, KEGG, and GO information were obtained from the Phytozome database [25]. The collection of CRCs can be found in Supplementary Table 3. Protein-protein interaction information was sourced from the BioGRID database [36]. The expression values of 35 maize Snf2 genes in various tissues at different developmental stages were obtained from MaizeMine database [37]. The expression values of 41 *Arabidopsis* Snf2 genes at different developmental stages were obtained from EBI's Expression Atlas [38].

2.2. Identification of Snf2 proteins

A total of 310 Snf2 proteins were collected, belonging to 24 Snf2

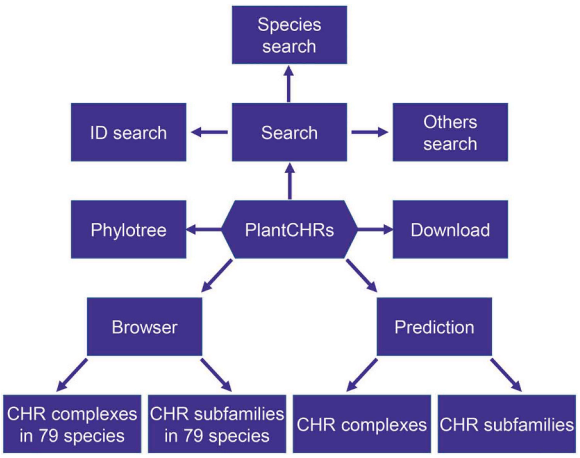

Fig. 1. The structure of the PlantCHRs. PlantCHRs are comprised of five modules: Search, Browse, Phylotree, Prediction, and Download. The Search module allows users to find relevant CHR proteins and CRCs through "ID search," "species search," and "other keyword search." The Browse module and Phylotree module enables users to explore CHR proteins and CRCs from 79 species in PlantCHRs categorized by species. The Prediction module allows users to predict CHR proteins and CRCs in other species.

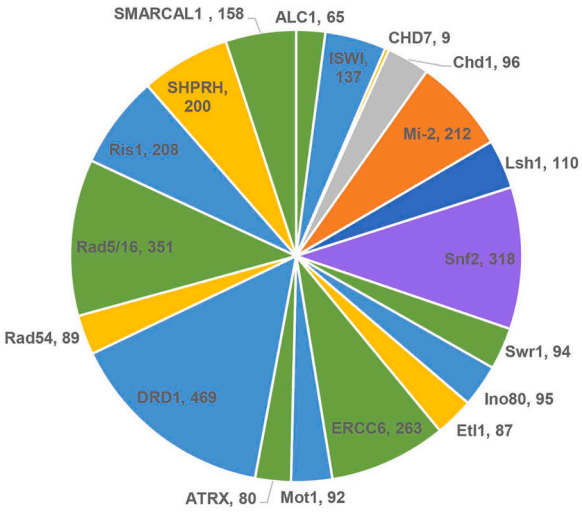

Fig. 2. The distribution of Snf2 genes across different subfamilies in 79 plant species was represented using a pie chart, which was divided into 19 sections, each indicating the number of genes in different Snf2 subfamilies. The labels on each section provided the name of the subfamily and the number of genes in that particular subfamily.

Table 2
Statistics of data in PlantCHRs.

Content	Number	Source
Species	79	
CHRs	3135	In house Prediction
CRC subunits	8398	In house Prediction
CDS sequence	11,533	PhytozomeV12, V13
Transcript sequence	11,533	PhytozomeV12, V13
Protein sequence	11,533	PhytozomeV12, V13
Protein domain	69,363	In house Prediction
PPIs	64,238	In house Prediction
GO term	33 (5362)	PhytozomeV12, V13
KEGG term	44 (4256)	PhytozomeV12, V13
Best hit with NR, Swissprot, Refseq databases	41,520	In house Prediction

subfamilies (Table 1). To establish HMM profiles for these subfamilies, multiple sequence alignment analysis was performed using MUSCLE v3.8.1551 software [39] followed by manual removal of poorly matched sequence alignment results. Then, HMMER v3.3.2 software [40] was used to create the HMM profiles. To assess the predictive performance of the HMM profiles, ROC curves were generated, and cutoff values for sequence scores and domain scores for each subfamily were calculated. This process involved two steps:

(1) Generation of the test data set: Snf2 subfamily members in several distantly related species were identified using sequence alignment analysis, protein domain identification, and phylogenetic tree construction. A total of 310 well-annotated Snf2 protein sequences were used as query sequences, and BLASTP software was used to align them against proteome sequences of *Physcomitrium patens*, *Populus*, *Ginkgo biloba*, *Gnetum montanum*, and *Amborella trichopoda* with e-value to $\leq$ 1e-10. Then, protein domains of aligned sequences were identified using InterProScan software [41]. Snf2 proteins with Pfam domains PF00176 (SNF2-related domain) and PF00271 (Helicase conserved C-terminal domain) were considered as Snf2 proteins. These proteins were aligned with the 310 well-annotated Snf2 protein sequences, and a phylogenetic tree was constructed using PHYLIP [42]. The subfamilies of newly predicted Snf2 proteins were determined based on the branching information in the phylogenetic tree (Supplementary Figure 7-11, Supplementary Table 4).

(2) Drawing of ROC curves: Based on the subfamily HMM profile, HMMER v3.3.2 software was used to scan the newly predicted Snf2 proteins, calculating the sequence scores and domain scores of all proteins. For each subfamily of the test set, members of the specific subfamily were considered positive data, while members of other subfamilies were considered negative data. Specificity and sensitivity values of the HMM profile under different sequence scores and domain scores were calculated, ROC curves were drawn, and AUC values and Youden index were calculated. The sequence scores and domain scores were used as thresholds for the HMM profile when the Youden index was the largest. The AUC values of all subfamily HMM profiles were above 0.99, indicating that the HMM profile has a strong discriminative ability (Supplementary Figure 1-6).

$$Sensitivity = \frac{TP}{TP + FN}$$

$$Specificity = \frac{TN}{TN + FP}$$

$$Youden \quad index = Sensitivity + Specificity - 1$$

In conclusion, by following this methodology, Snf2 proteins in various species were identified, and HMM profiles for Snf2 subfamilies with high discriminative ability were established. This approach allows for the accurate prediction of Snf2 proteins in other plant species, contributing to the understanding of the roles of these proteins in plant growth, development, and stress response.

Based on the HMM profiles of the Snf2 subfamilies (ALC1, Arip4, ATRX, Chd1, CHD7, DRD1, ERCC6, Etl1, Ino80, ISWI, JBP2, Lodestar, Lsh1, Mi-2, Mot1, Rad5/16, Rad54, Ris1, SHPRH, SMARCAL1, Snf2, and Swr1), the HMMER software [40] was used to predict Snf2 subfamily proteins in new species. However, due to the limited number of members in the EP400 and SSO1653 subfamilies (fewer than three each), they could not meet the requirements for HMM profile construction. To identify potential EP400 and SSO1653 subfamily members in these species, the collected EP400 and SSO1653 subfamily members were used as query sequences, and the proteome sequences of the species to be identified were used as the subject database. BLASTP software was used for alignment. Subsequently, the presence of the PF00176 and

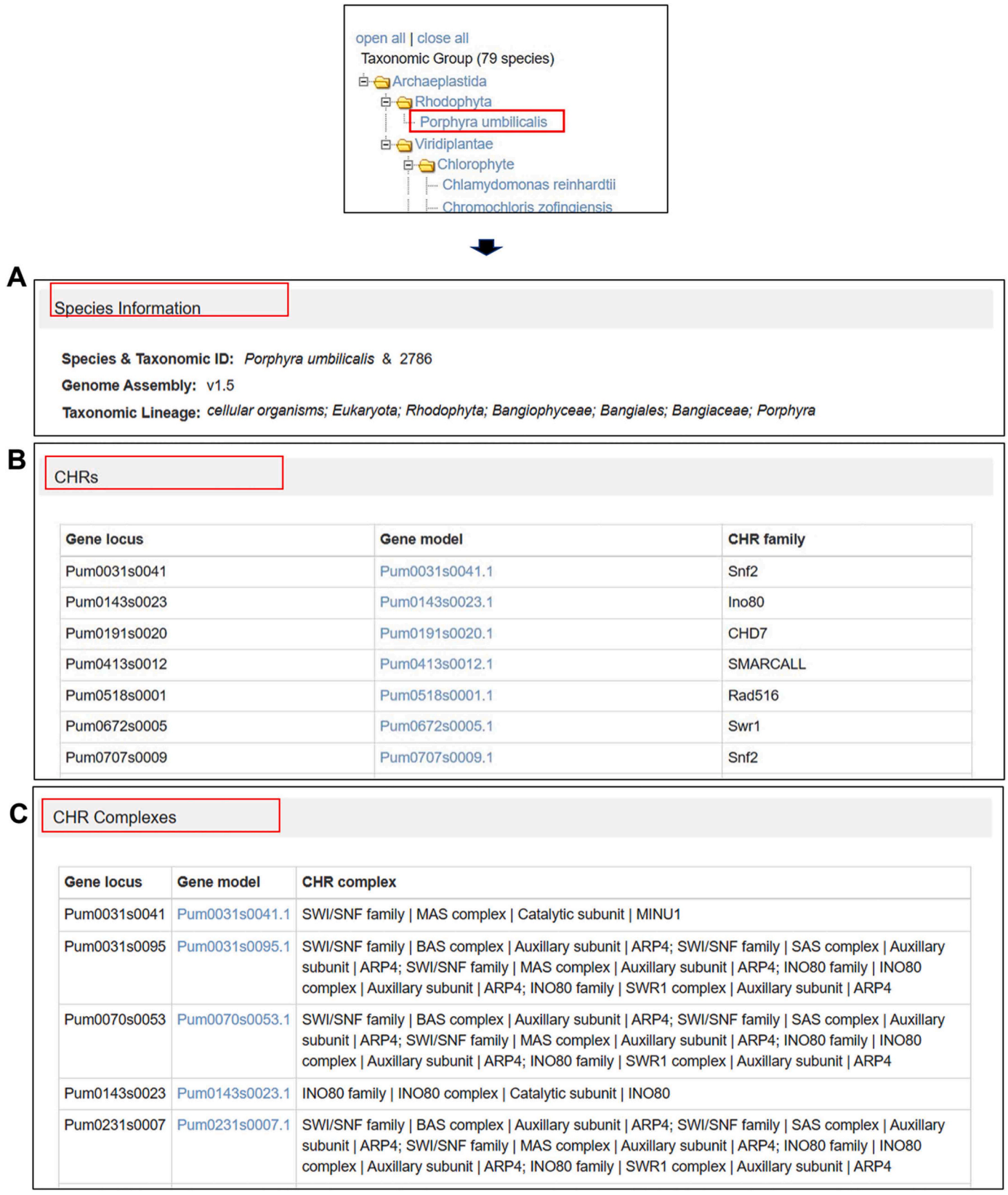

Fig. 3. Browse page of PlantCHRs.

PF00271 domains was detected in the aligned proteins. Moreover, it was necessary to exclude other Snf2 proteins identified using the HMM profile method to obtain EP400 and SSO1653 subfamily members.

2.3. Identification of CHR complex proteins

CHR complex proteins were predicted using the homologous protein identification method. Using the 84 collected Arabidopsis thaliana CHR

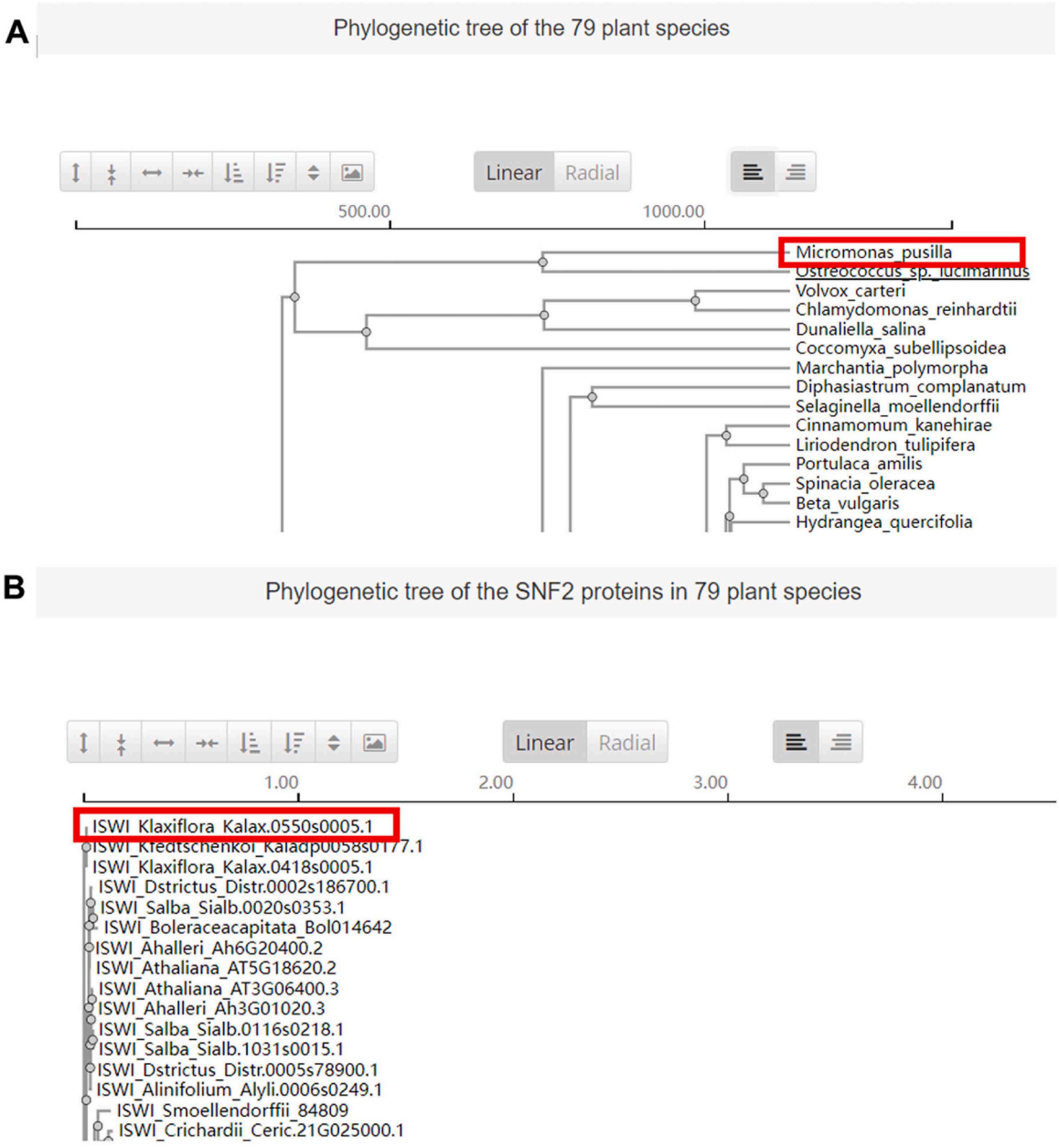

Fig. 4. Phylotree page of PlantCHRs.

complex proteins as query sequences (Supplementary Table 3), the Inparanoid software [43] was employed to predict CHR complex proteins in other species.

2.4. Protein-protein interaction

With the help of protein-protein interaction pairs of *Arabidopsis thaliana* obtained from Biogrid, we used OrthoFinder [44] to predict homologous proteins of these interacting proteins in other species, thereby predicting protein-protein interaction in other species.

2.5. Protein domain identification

Locally InterProScan software [41] was used to identify protein domains (CDD, Pfam, SMART), gene family classifications (PANTHER, SUPERFAMILY), and other information of protein sequences. The E-value was set as ≤ 0.001.

2.6. Protein annotation

The proteins were aligned with the proteins of Uniprot, Swiss-Prot, TrEMBL, NR database, and the *Arabidopsis thaliana* protein sequence. The best-hit protein was selected for functional annotation.

2.7. Prediction of physicochemical properties of proteins

Protein Calculator Parameter tool in TBtools [45] was used to predict the relative molecular weight (kDa), isoelectric point (PI), instability index, grand average of hydropathicity (GRAVY), and aliphatic index of the proteins.

2.8. Implementation

Based on our data collection and functional annotation, we developed the Plant Chromatin Remodeling Factors Database (PlantCHRs) using the classic LAMP (Linux + Apache + MySQL + PHP) environment.

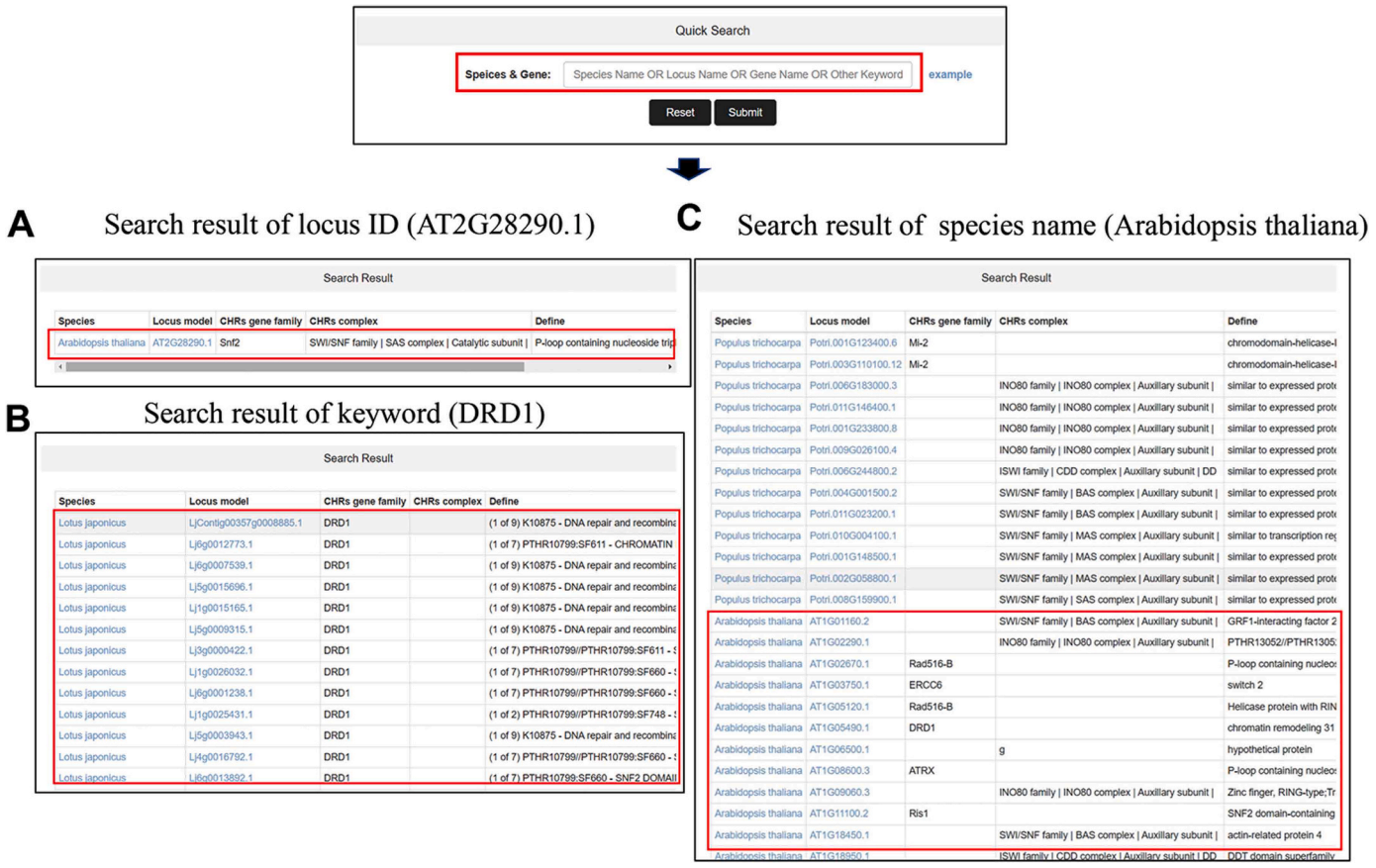

Fig. 5. Search page of PlantCHRs.

PlantCHRs comprises five modules: search, browse, phylotree, prediction, and download (Fig. 1). The search tools include gene ID search, species name search, and keyword search. Through the browsing tool, users can explore all CHR proteins and CRCs in species classification manner. The amino acid sequence of Snf2 proteins from 79 species within PlantCHRs was used to construct a phylogenetic tree, which is accessible through the phylotree module. Additionally, the time tree of the 79 species was integrated phylotree module, which was constructed using the Timetree webserver. In the prediction section, users can use this tool to predict unknown CHR proteins and CHR complex proteins. The download page provides sequence information and gene family downloads for CHR proteins and CHR complex proteins of different species.

3. Result

3.1. Statics of PlantCHRs

PlantCHRs contains a total of 3135 CHRs and 8398 CRC subunits from 79 plants. These CHRs are classified into 19 subfamilies, with DRD1, Snf2, Rad5/16, ERCC6, and Mi-2 being the subfamilies with a significantly higher number of genes. However, some subfamilies have relatively few genes, such as CHD7, which only has 9 members (Fig. 2). Representative CDS sequences, transcript sequences, and protein sequences of these factors were also included in PlantCHRs (Table 2). To facilitate studies of these factors, GO and KEGG terms, PPIs, and best hit with other protein databases were also integrated into PlantCHRs (Table 2).

3.2. Usage

PlantCHRs comprises CHRs obtained from a wide range of evolutionarily significant plant species, including red algae, green algae, club moss, ferns, gymnosperms, and angiosperms. To facilitate user access to the data in PlantCHRs, multiple options are provided, namely: browsing, phylotree, and searching. In the browsing module, users can explore CHRs and CRCs organized by species classification. The resulting page is divided into three sections: species information, CHRs, and CRCs (Fig. 3). In the phylotree module, we employed the Timetree web server to construct an evolutionary tree representing 79 plant species. Users can utilize this phylogenetic tree to search for CHRs of different species (Fig. 4A). Furthermore, we performed a phylogenetic analysis of Snf2 proteins across the 79 plant species and constructed a corresponding phylogenetic tree. Users can search specific branches of Snf2 proteins by the phylogenetic tree (Fig. 4B). In the search module, users can search for locus IDs, keywords, and species names. After entering the relevant search terms and clicking the search button. The concise information on the search results will be presented on the search results page (Fig. 5). In addition, we have designed a tool that can predict CHRs and CRCs in newly sequenced species. After the user uploads the protein sequence and submits it, the tool can determine whether CHRs and CRCs are present in the submitted protein sequence. By using this tool, we predicted the full set of Snf2 proteins encoded by the barley genome (Fig. 6), and a total of 39 proteins were predicted as CHRs, consistent with the results predicted by Chen et al., which indicated the reliability of PlantCHRs' predictive results (Table 3) [46].

We exemplified the gene detail page of PlantCHRs with the tomato Snf2 gene (*Solyc01g109970*). The search results are organized into seven sections: basic information, family classification, protein domain, GO, KEGG, protein-protein interactions, and protein annotation (Fig. 7). In

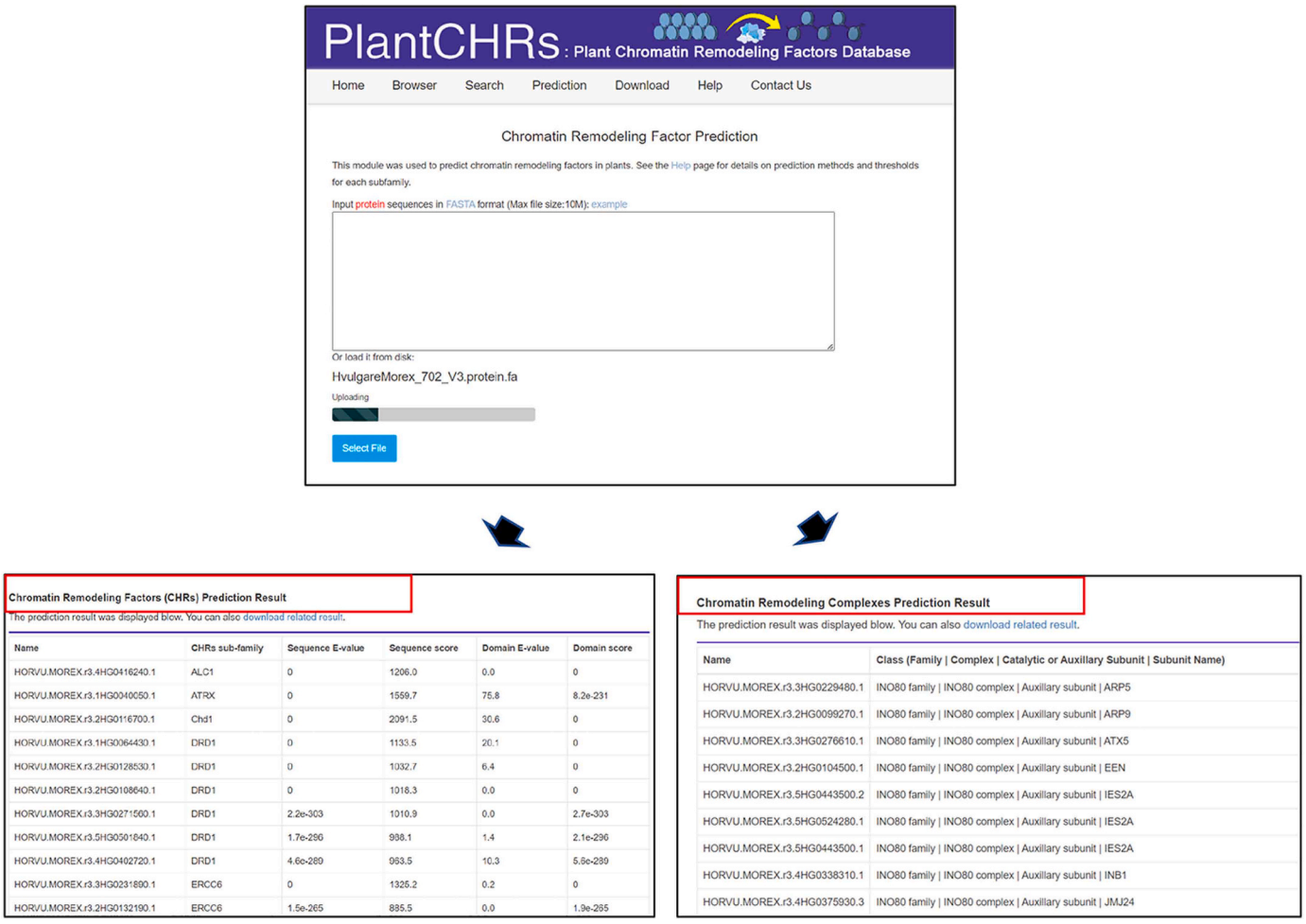

Fig. 6. Prediction page of PlantCHRs.

the basic information section, users can acquire information about the Snf2 subfamily to which the gene belongs, the species it belongs to, and protein properties, including its molecular weight, theoretical pI, Instability Index, Aliphatic Index, and GRAVY. Moreover, users can download the CDS, transcript, and protein sequences of the gene (Fig. 7A). The family classification section displays the results of the protein alignment with the DRD1 hmm profile (Fig. 7B). The protein domain section lists the annotations and gene family classification information from various sources, including cd18793 (SF2_C_SNF), PF00271 (Helicase conserved C-terminal domain), PF00176 (SNF2-related domain), SM00490 (helicmild6), SM00487 (ultradead3), and PTHR45821 (SNF2 DOMAIN-CONTAINING PROTEIN CLASSY 2-RELATED). These pieces of information indicate that Solyc01g109970 belongs to the Snf2 family (Fig. 7C). In the GO and KEGG sections, users can learn that the protein has ATP-binding ability and participates in the homologous recombination process (Fig. 7D, E). The protein-protein interaction section reveals that Solyc01g109970 has a series of interacting proteins such as Solyc01g096390 (nuclear RNA polymerase D1B), Solyc01g098320 (Eukaryotic rpb5 RNA polymerase subunit family protein), Solyc05g010300 (DNA-directed RNA polymerase family protein), Solyc03g071790 (defective in meristem silencing 3), Solyc09g082420 (RNA-DIRECTED DNA METHYLATION 1), and Solyc04g040170 (Eukaryotic rpb5 RNA polymerase subunit family protein). These proteins provide valuable insight into the molecular mechanisms of Solyc01g109970's involvement in biological processes (Fig. 7F). Lastly, the protein annotation section lists the best hit protein of Solyc01g109970 in Refseq, Swissprot, TrEMBL, and TAIR databases, which can broaden users' understanding of the annotation of *Solyc01g109970* (Fig. 7G).

3.3. Comparative analysis of the distribution of Snf2 members in different species

Using the data from PlantCHRs, we performed a comparative analysis of Snf2 proteins in 37 plant species, 13 animal species, and 3 fungal species, significant differences were found in the total number and subfamily types of Snf2 genes among different species during the evolutionary process. For instance, ascomycete fungi possess no more than 10 types of Snf2 subfamilies, comprising a total gene count of no more than 14. Nematodes and insects possess 11–13 subfamilies, consisting of 13 Snf2 genes. Red algae and green algae have 7 and approximately 13 Snf2 gene families, respectively. However, most chordates and flowering plants (86.5%) possess over 17 Snf2 subfamily types and more than 30 genes (Fig. 8 and Supplementary Table 5). These observations indicate that the total number and subfamily types of Snf2 proteins in fungi, lower animal and plant are significantly lower than those in higher animal and plant. Differences also exist in Snf2 subfamily types among various species. For instance, ALC1, Lsh1, EP400, Arip4, DRD1, Lodestar, and SMARCAL1 are absent in fungi. All animals possess the CHD7 subfamily, and only some lower plants, such as mosses and ferns, have the CHD7 subfamily. Lodestar is exclusive to animals, Arip4 and EP400 are unique to chordates, and DRD1 is found in all green plants except green algae. Ris1 is detected in most plants but not in any animals. SSO1653 is identified in only one plant, and JBP2 is absent in all species (Fig. 8 and Supplementary Table 5).

Our study also found that the number of genes within specific Snf2

Table 3
Prediction result of Snf2 genes in Barley by Chen et al. and PlantCHRs.

Locus ID	Predicted by Chen et al.	PlantCHRs Prediction
HORVU.MOREX.r3.7HG0669610	Snf2	Snf2
HORVU.MOREX.r3.6HG0543700	Snf2	Snf2
HORVU.MOREX.r3.1HG0019730	Snf2	Snf2
HORVU.MOREX.r3.2HG0184370	Lsh	Lsh1
HORVU.MOREX.r3.4HG0338270	Lsh	Lsh1
HORVU.MOREX.r3.7HG0646770	Lsh	Lsh1
HORVU.MOREX.r3.1HG0022440	Iswi	ISWI
HORVU.MOREX.r3.3HG0230070	Iswi	ISWI
HORVU.MOREX.r3.4HG0416240	Iswi	ALC1
HORVU.MOREX.r3.2HG0116700	Iswi	Chd1
HORVU.MOREX.r3.7HG0658830	Mi-2	Mi-2
HORVU.MOREX.r3.3HG0307470	Mi-2	Mi-2
HORVU.MOREX.r3.2HG0143700	Mi-2	Mi-2
HORVU.MOREX.r3.7HG0748860	Swr1	Swr1
HORVU.MOREX.r3.4HG0375120	Ino80	Ino80
HORVU.MOREX.r3.2HG0187800	Etl1	Etl1
HORVU.MOREX.r3.6HG0617770	Rad54	Rad54
HORVU.MOREX.r3.1HG0040050	ATRX	ATRX
HORVU.MOREX.r3.2HG0108640	DRD1	DRD1
HORVU.MOREX.r3.4HG0402720	DRD1	DRD1
HORVU.MOREX.r3.3HG0271560	DRD1	DRD1
HORVU.MOREX.r3.1HG0064430	DRD1	DRD1
HORVU.MOREX.r3.2HG0128530	DRD1	DRD1
HORVU.MOREX.r3.5HG0501840	DRD1	DRD1
HORVU.MOREX.r3.2HG0119710	Rad5/16	Rad5/16
HORVU.MOREX.r3.6HG0559070	Rad5/16	Rad5/16
HORVU.MOREX.r3.2HG0103780	Rad5/16	Rad5/16
HORVU.MOREX.r3.2HG0141740	Rad5/16	Rad5/16
HORVU.MOREX.r3.3HG0293510	Ris1	Ris1
HORVU.MOREX.r3.2HG0199060	Ris1	Ris1
HORVU.MOREX.r3.7HG0696030	Ris1	Ris1
HORVU.MOREX.r3.3HG0320710	SHPRH	SHPRH
HORVU.MOREX.r3.1HG0094930	Mot1	Mot1
HORVU.MOREX.r3.3HG0231890	ERCC6	ERCC6
HORVU.MOREX.r3.2HG0132190	ERCC6	ERCC6
HORVU.MOREX.r3.2HG0217790	ERCC6	ERCC6
HORVU.MOREX.r3.6HG0559010	ERCC6	ERCC6
HORVU.MOREX.r3.2HG0128410	SMARCALL	SMARCALL
HORVU.MOREX.r3.UnG0786740	Lsh1	Lsh1

subfamilies is generally consistent among different animals. However, variations exist in the number of genes within specific Snf2 subfamilies among different plant species. Subfamilies with more notable variations in the number of Snf2 genes between plant species include Snf2, DRD1, Rad5/16, and Ris1 (Fig. 8 and Supplementary Table 5). These results imply that the functions of Snf2 genes may differ between plants and animals.

3.4. Phylogenetic and protein domain analyses of Snf2 family genes in four representative species

We performed phylogenetic and domain analyses of Snf2 proteins from four representative species: *A. thaliana* (dicot plant), *Amborella trichopoda* (basal angiosperm), *Ginkgo bilobas* (Gymnospermae species), and *Physcomitrella patens* (Bryophyta species) to investigate domain variations among different Snf2 subfamilies. Snf2 proteins from these four representative plants were classified into 19 branches, each representing a distinct Snf2 subfamily (Fig. 9). All Snf2 proteins contained the "Helicase conserved C-terminal domain" and "SNF2-related domain". The "Chromo domain" was exclusive to the CHD7, Chd1, and Mi-2 subfamilies. The ISWI subfamily exclusively contained the "HAND," "SANT," and "SLIDE" domains. The Mi-2 subfamily exclusively contained the "CHD subfamily II, DUF1087" and "PHD-finger" domains, and some Mi-2 subfamily proteins lacked "CHD subfamily II, DUF1087". The "Ring-finger" domain and "HIRAN domain" were present in the Rad5/16 and Ris1 subfamilies. The Rad5/16 subfamily exclusively contained the "HIRAN domain". The Swr1 subfamily exclusively contained the "HSA domain". The Ino80 subfamily exclusively contained the "DNA-binding domain". Additionally, we discovered additional subdivisions within certain subfamilies. Within one branch of the Snf2 subfamily, the C-terminus of the protein corresponds to the "Snf2-ATP coupling, chromatin remodeling complex," whereas in another branch, the C-terminus of Snf2 corresponds to the "Bromodomain". The SMARCAL1 subfamily can be further divided into two branches: one branch with the C-terminus corresponding to the "Histone H1-like protein Hc1," and the other branch lacking this domain.

3.5. Analysis of expression patterns of maize and Arabidopsis Snf2 genes in different tissues

We also obtained Snf2 genes of *Arabidopsis* and maize from PlantCHRs and investigated their expression patterns during growth and development. The expression of Snf2 genes in various tissues of maize and *Arabidopsis* exhibited a certain degree of conservation. The majority of Snf2 genes in both species exhibit relatively high expression levels in developing flower and meristem tissues, including immature tassels, immature cobs, meiotic tassels, stem and shoot apical meristems, elongated internodes in maize, and flowers and callus in *Arabidopsis* (Fig. 10). Additionally, in both maize and *Arabidopsis*, a subset of genes displays lower expression levels in leaf tissues, including four maize *DRD1* genes, five *Arabidopsis DRD1* genes, three maize *Rad5/16* genes, four *Arabidopsis Rad5/16* genes, two maize *Lsh1* genes, and one *Arabidopsis Lsh1* gene (Fig. 10). Notably, the expression of Snf2 genes in maize and *Arabidopsis* also exhibits some discrepancies. For example, numerous Snf2 genes display higher expression levels in maize seeds, whereas the expression of several Snf2 genes in *Arabidopsis* seeds is

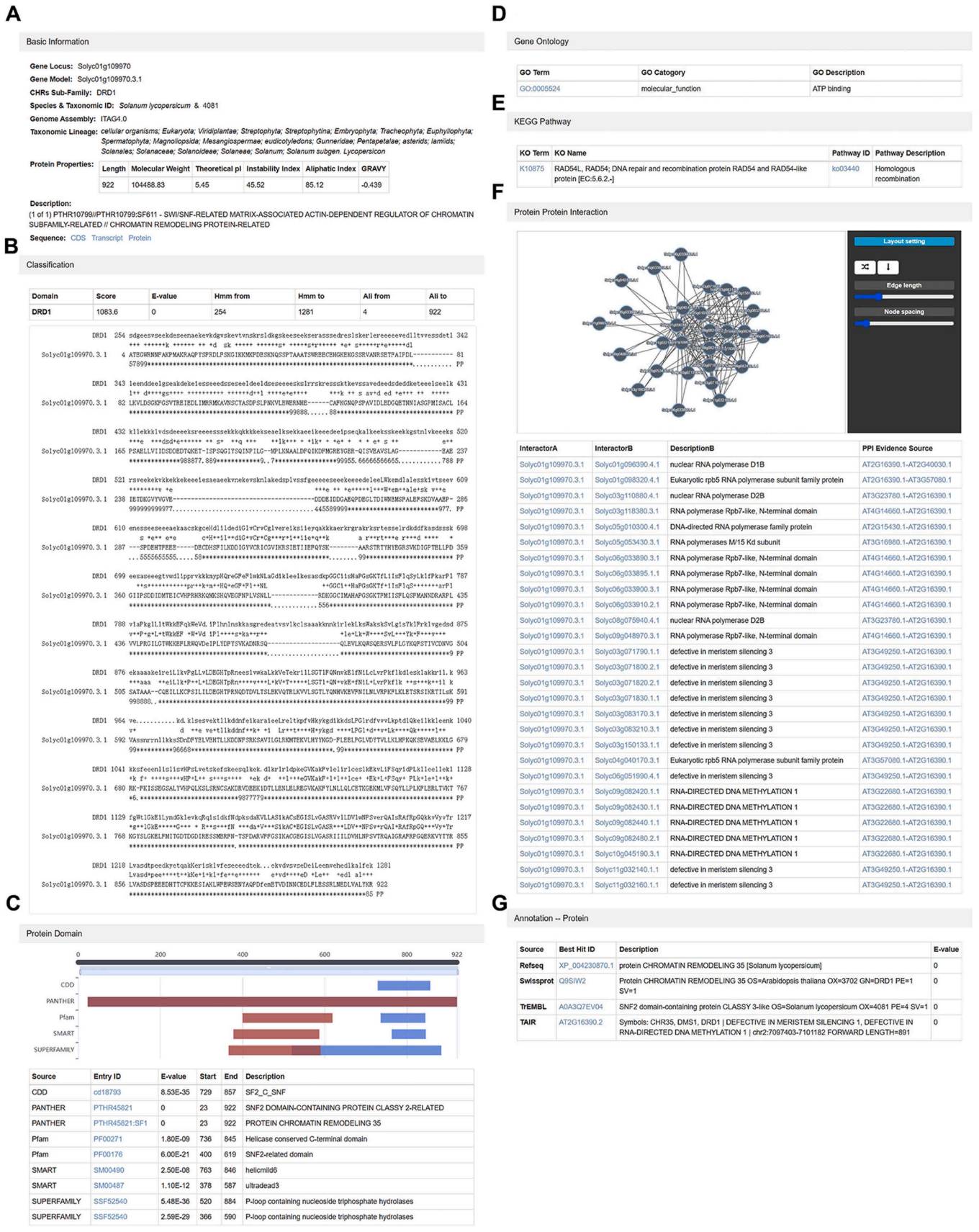

Fig. 7. The gene detail page of Tomoto CHR gene (*Solyc01g109970*), which includes seven sections: (A) Basic information, including gene locus, family, protein properties, description, and sequence. (B) Classification, including alignment results with HMM model of CHR subfamily. (C) Protein domain. (D) GO information. (E) KEGG information. (F) PPI information. (G) Annotation information, including the best hit of the protein in Refseq, Swissprot, TrEMBL, and TAIR databases.

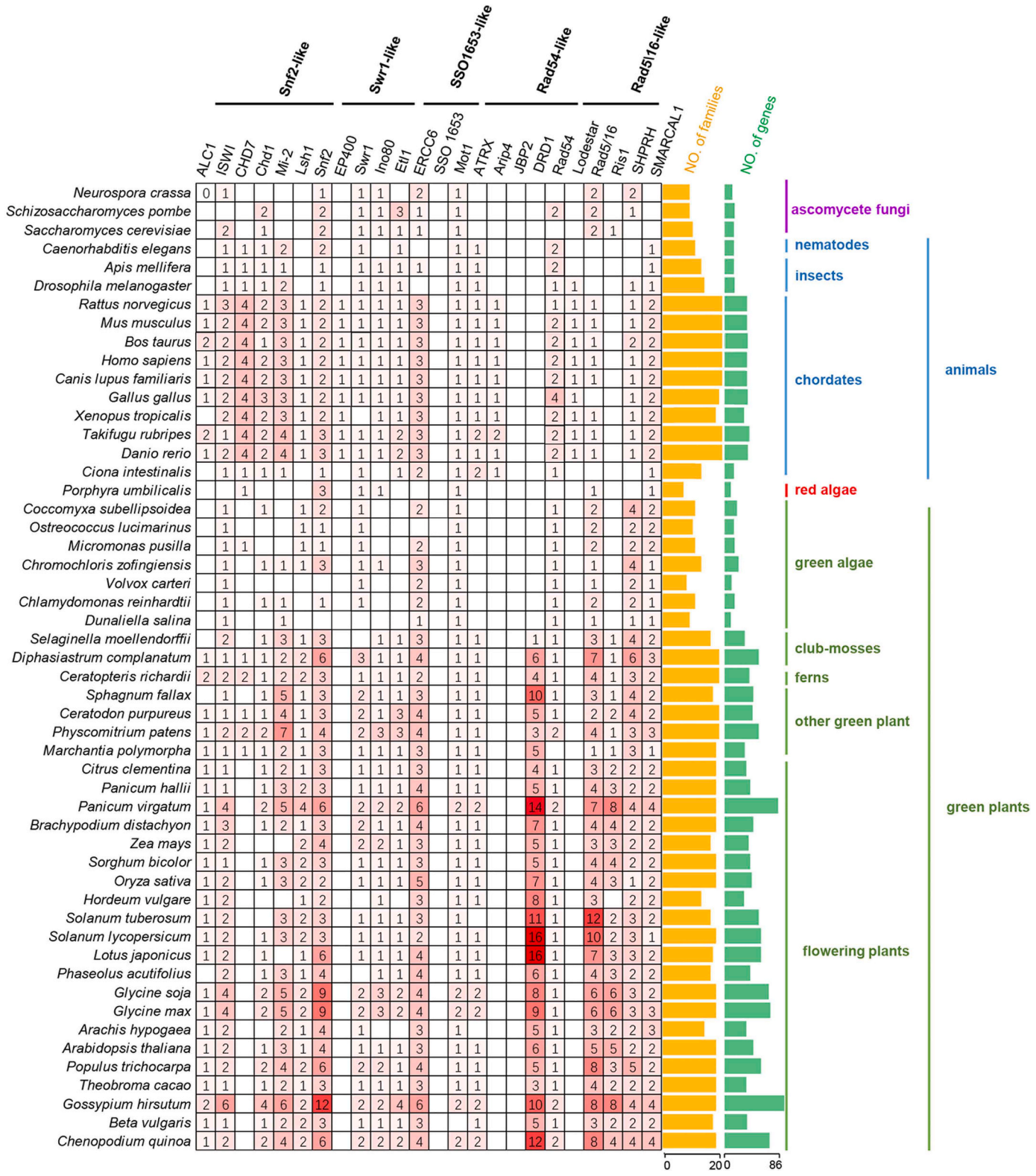

Fig. 8. The number of SNF2 subfamilies in different species. The horizontal axis represents different SNF2 subfamilies, the left vertical axis represents different species, and the right vertical axis represents different stages during evolution. In the heatmap, each number in the cell represents the number of SNF2 proteins, with darker red color indicating higher number. The yellow and green bar charts on the right of the heatmap represent the distribution of SNF2 subfamily and gene numbers in different species.

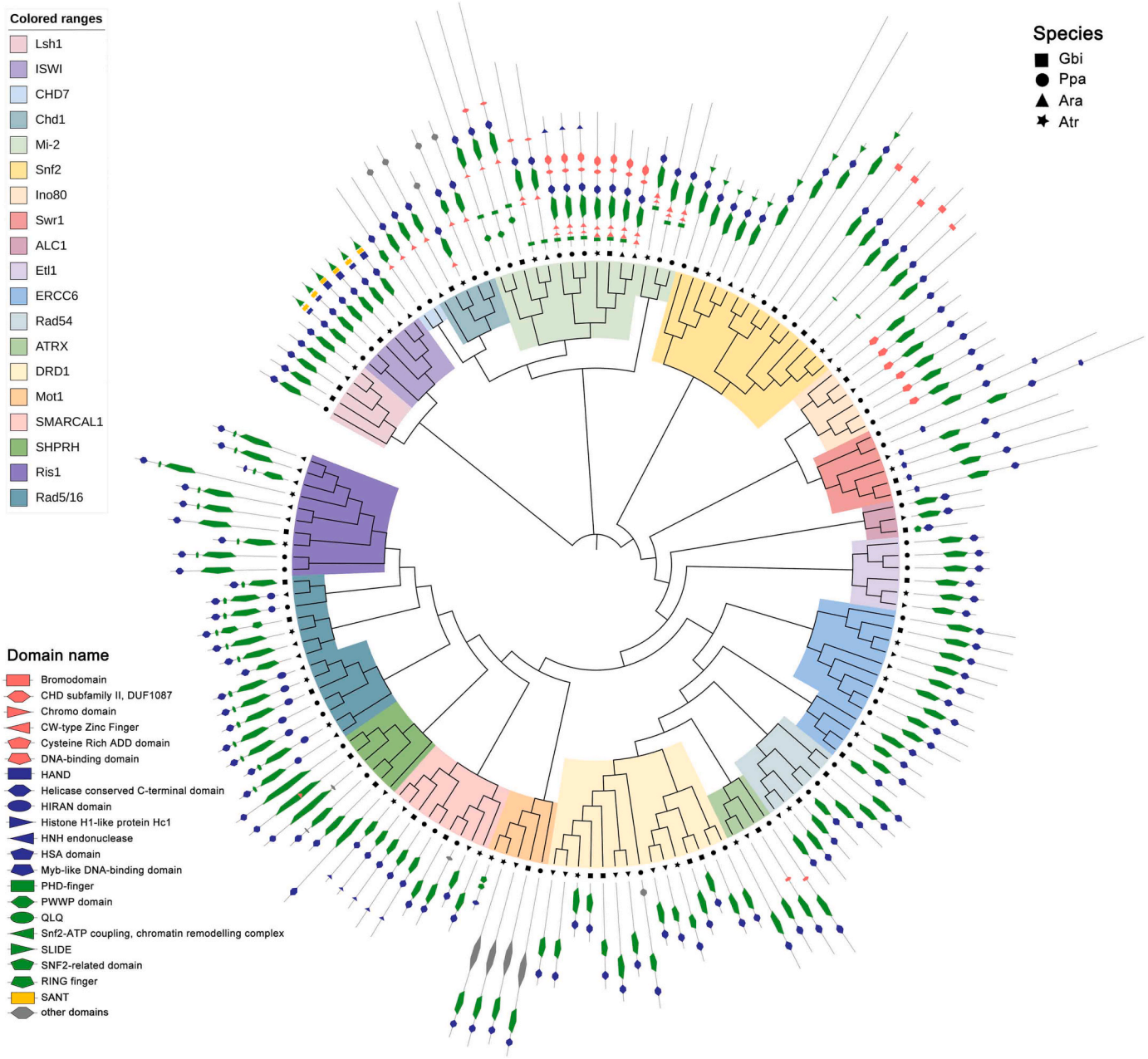

Fig. 9. Protein domain analysis of Snf2 proteins in four representative plant, namely *Arabidopsis thaliana, Ginkgo biloba, Physcomitrium patens*, and *Amborella trichopoda*. Maximum-likelihood trees were constructed based on the amino acid sequences of the Snf2 gene family. The identification of protein domains was performed using the InterProScan program. In this study, the abbreviations "Ath", "Atr", "Gbi", and "Ppa" were used to represent "*Arabidopsis thaliana*", "*Amborella trichopoda*", "*Ginkgo biloba*", and "*Physcomitrium patens*", respectively.

comparatively low (Fig. 10).

4. Discussion and conclusion

The Snf2 protein functions as the catalytic subunit of chromatin remodeling complexes, dynamically regulating gene expression by altering the nucleosome position on chromatin. As sessile organisms, plants require chromatin remodeling factors to modulate the expression of related genes, enabling adaptation to diverse unfavorable environments during growth [47]. Currently, our knowledge of plant CHRs primarily derives from studies in the model plant *Arabidopsis thaliana* and a few crop species; however, the role of CHRs in most plants remains elusive [5,48,35,34]. In this study, we generated HMM profiles of various Snf2 subfamilies using 310 well-annotated Snf2 proteins. By creating test datasets, we demonstrated the reliability of these HMM profiles. Using these profiles, we identified Snf2 proteins in 79 plant species, uncovering a total of 3135 Snf2 proteins belonging to 19 subfamilies. Snf2 proteins typically function as catalytic subunits of CRCs, we also predicted 8398 CRC protein subunits in 79 plants.

Cross-species comparative analysis reveals that Snf2 proteins are crucial in plant evolution. Various adverse factors in the natural environment can threaten plant survival. Numerous studies have shown that Snf2 proteins are involved in plant growth and stress response [13]. Consequently, during natural selection, plants may increase the copy number and subfamily of Snf2 proteins to adapt to the ever-changing environment, enhancing their stress resistance and facilitating their survival. Significant expansion of Snf2 proteins and their subfamilies is observed in the evolutionary stages of mosses/ferns, suggesting that

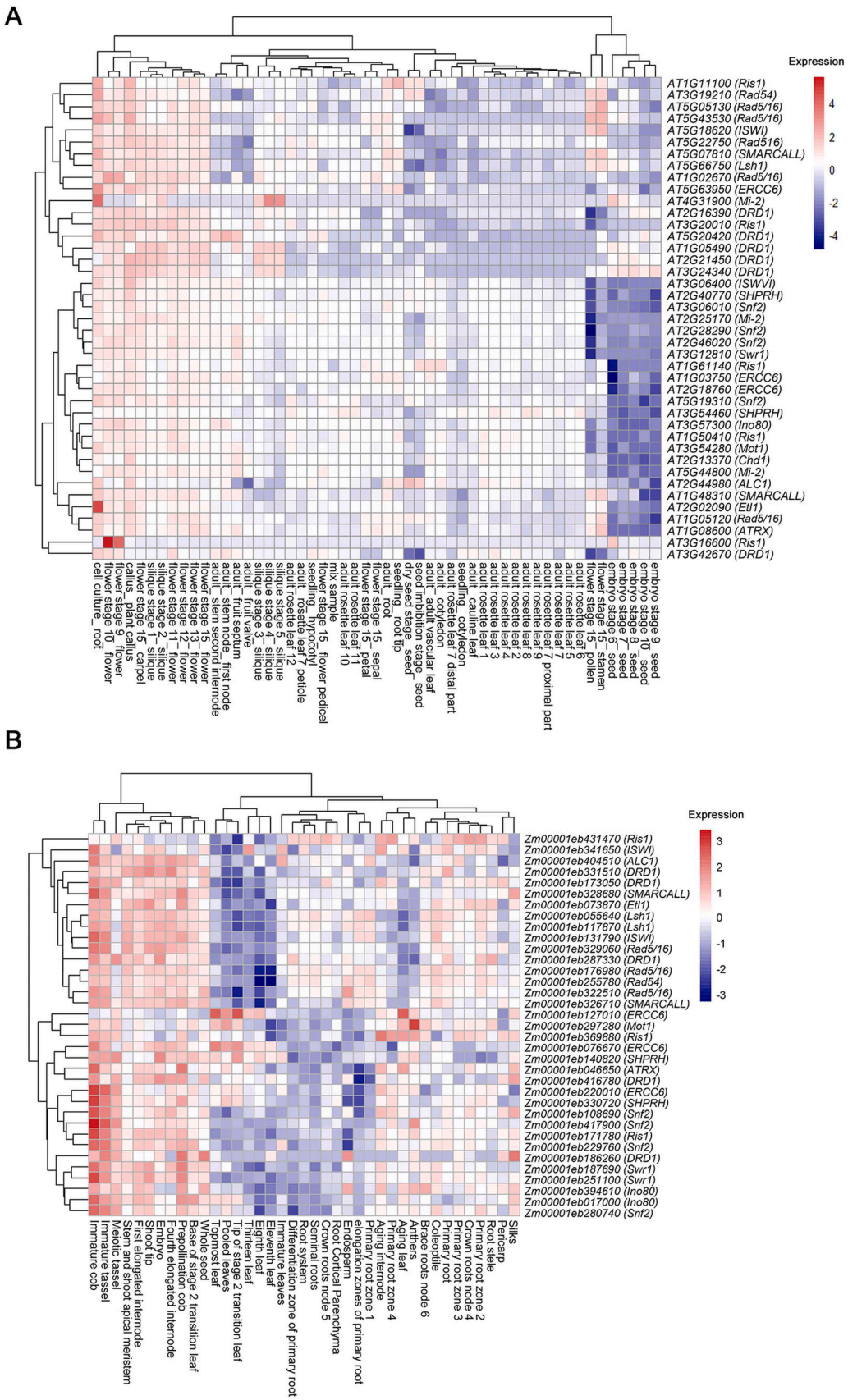

Fig. 10. Heatmap show the expression patterns of CHR genes in different tissues and developmental stages of *Arabidopsis* and maize. The expression levels of the genes were represented by $\log_2$(FPKM) which was normalized by Z-score. The color of each cell in the heatmap corresponds to the expression level of the gene, where red indicates high expression and blue indicates low expression.

Snf2 proteins may contributed to the evolution of plants from aquatic to terrestrial life (Fig. 8). Furthermore, our findings suggest that animals possess fewer Snf2 proteins than plants, which might be related to plants being sessile organisms; they encounter more adverse environmental factors during their growth than animals. Unique Snf2 gene families are present in both animals and plants, such as Arip4, Lodestar, and EP400 in animals, and DRD1 and Ris1 in plants. DRD1 exhibits an increased copy number in several plant, including *Panicum virgatum*, *Brachypodium mexicanum*, *Solanum tuberosum*, *Solanum lycopersicum*, *Lotus japonicus*, *Salix purpurea*, *and Chenopodium quinoa*. The significance of species-specific expansion, however, remains unclear and warrants further investigation. Moreover, the identification of plant-specific subunits and their functions could provide valuable insights into the unique aspects of plant chromatin regulation and the adaptation mechanisms employed by plant in response to environmental challenges.

In summary, we have established an online database for plant chromatin remodeling factors (PlantCHRs), which encompasses CHRs and CRCs of 79 plants. To facilitate the function analysis of these CHRs and CRCs, we have incorporated annotations, such as: protein domains, GO, KEGG, and PPI information. Furthermore, the PlantCHRs database will be continuously updated and improved. We will update the CHRs of newly sequenced species into PlantCHRs, and collect information on other epigenetic modification factors, providing a more comprehensive database for epigenetic factors in plants.

Funding

This research was funded by Talent Introduction Special Funds of Qingdao Agricultural University, grant number: 663/1121029; Shandong Province Key Research and Development Plan (Major Science and Technology Innovation Project), grant number: 2021LZGC013; the supply chain integration and information platform design of QINGDAO HUALUTE MACHINE CO., LTD, grant number: 6602423707; the Qingdao Agricultural University High-Level Talent Special Support Program, grant number: 665/1120002; the 2021 Qingdao Municipal Science and Technology Plan Project: Science and Technology Benefiting the People Demonstration and Guidance Special Support Project, grant number: 21–1–4-ny-23-nsh; Blueberry Big Data and Integrated Water-Fertilizer Technology, grant number: 2021TZXD003–002; the 2020 Research Program of Sanya Yazhou Bay Science and Technology City, grant number: 2020-02-005.

CRediT authorship contribution statement

Conceptualization, H.Y. and H.C.; methodology, H.Y. and J.Y.; software, G.Z., S.L., W.M. and T.Y.; formal analysis, H.Y. and G.Z.; data curation, H.Y., F.L., S.L., W.M. and T.Y.; writing - original draft, H.Y. and F.L.; writing - review & editing, H.Y., F.L. and Y.L.; funding acquisition, H.Y., Y.L. and H.C. All authors have read and agreed to the published version of the manuscript.

Conflict of Interest

The authors declare no conflict of interest.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.csbj.2023.10.005.

References

[1] Blosser TR, Yang JG, Stone MD, Narlikar GJ, Zhuang X. Dynamics of nucleosome remodelling by individual ACF complexes. Nature 2009;462:1022–7.
[2] Harada BT, Hwang WL, Deindl S, Chatterjee N, Bartholomew B, Zhuang X. Stepwise nucleosome translocation by RSC remodeling complexes. Elife 2016:5.
[3] Willhoft O, Ghoneim M, Lin CL, Chua EYD, Wilkinson M, Chaban Y, Ayala R, McCormack EA, Ocloo L, Rueda DS, et al. Structure and dynamics of the yeast SWR1-nucleosome complex. Science 2018:*362*.
[4] Durr H, Flaus A, Owen-Hughes T, Hopfner KP. Snf2 family ATPases and DExx box helicases: differences and unifying concepts from high-resolution crystal structures. Nucleic Acids Res 2006;34:4160–7.
[5] Flaus A, Martin DM, Barton GJ, Owen-Hughes T. Identification of multiple distinct Snf2 subfamilies with conserved structural motifs. Nucleic Acids Res 2006;34:2887–905.
[6] Knizewski L, Ginalski K, Jerzmanowski A. Snf2 proteins in plants: gene silencing and beyond. Trends Plant Sci 2008;13:557–65.
[7] Farrona S, Hurtado L, Bowman JL, Reyes JC. The Arabidopsis thaliana SNF2 homolog AtBRM controls shoot development and flowering. Development 2004;131:4965–75.
[8] Han SK, Sang Y, Rodrigues A, Biol F, Wu MF, Rodriguez PL, Wagner D. The SWI2/SNF2 chromatin remodeling ATPase BRAHMA represses abscisic acid responses in the absence of the stress stimulus in Arabidopsis. Plant Cell 2012;24:4892–906.
[9] Tang X, Hou A, Babu M, Nguyen V, Hurtado L, Lu Q, Reyes JC, Wang A, Keller WA, Harada JJ, et al. The Arabidopsis BRAHMA chromatin-remodeling ATPase is involved in repression of seed maturation genes in leaves. Plant Physiol 2008;147:1143–57.
[10] Sang Y, Silva-Ortega CO, Wu S, Yamaguchi N, Wu MF, Pfluger J, Gillmor CS, Gallagher KL, Wagner D. Mutations in two non-canonical Arabidopsis SWI2/SNF2 chromatin remodeling ATPases cause embryogenesis and stem cell maintenance defects. Plant J 2012;72:1000–14.
[11] Mlynarova L, Nap JP, Bisseling T. The SWI/SNF chromatin-remodeling gene AtCHR12 mediates temporary growth arrest in Arabidopsis thaliana upon perceiving environmental stress. Plant J 2007;51:874–85.
[12] Li G, Zhang J, Li J, Yang Z, Huang H, Xu L. Imitation Switch chromatin remodeling factors and their interacting RINGLET proteins act together in controlling the plant vegetative phase in Arabidopsis. Plant J 2012;72:261–70.
[13] Liu H, Li J, Xu Y, Hua J, Zou B. ISWI chromatin remodeling factors repress PAD4-mediated plant immune responses in Arabidopsis. Biochem Biophys Res Commun 2021;583:63–70.
[14] Jing Y, Guo Q, Lin R. The chromatin-remodeling factor PICKLE antagonizes polycomb repression of FT to promote flowering. Plant Physiol 2019;181:656–68.
[15] Lee S, Choi J, Park J, Hong CP, Choi D, Han S, Choi K, Roh TY, Hwang D, Hwang I. DDM1-mediated gene body DNA methylation is associated with inducible activation of defense-related genes in Arabidopsis. Genome Biol 2023;24:106.
[16] Kang H, Liu Y, Fan T, Ma J, Wu D, Heitz T, Shen WH, Zhu Y. Arabidopsis CHROMATIN REMODELING 19 acts as a transcriptional repressor and contributes to plant pathogen resistance. Plant Cell 2022;34:1100–16.
[17] Hernandez Sanchez-Rebato M, Bouatta AM, Gallego ME, White CI, Da Ines O. RAD54 is essential for RAD51-mediated repair of meiotic DSB in Arabidopsis. PLoS Genet 2021;17:e1008919.
[18] Deal RB, Topp CN, McKinney EC, Meagher RB. Repression of flowering in Arabidopsis requires activation of FLOWERING LOCUS C expression by the histone variant H2A.Z. Plant Cell 2007;19:74–83.
[19] Pardal AJ, Piquerez SJM, Dominguez-Ferreras A, Frungillo L, Mastorakis E, Reilly E, Latrasse D, Concia L, Gimenez-Ibanez S, Spoel SH, et al. Immunity onset alters plant chromatin and utilizes EDA16 to regulate oxidative homeostasis. PLoS Pathog 2021;17:e1009572.
[20] Meng X, Yu X, Wu Y, Kim DH, Nan N, Cong W, Wang S, Liu B, Xu ZY. Chromatin remodeling protein ZmCHB101 regulates nitrate-responsive gene expression in maize. Front Plant Sci 2020;11:52.
[21] Cho SH, Lee CH, Gi E, Yim Y, Koh HJ, Kang K, Paek NC. The rice rolled fine striped (RFS) CHD3/Mi-2 chromatin remodeling factor epigenetically regulates genes involved in oxidative stress responses during leaf development. Front Plant Sci 2018;9:364.
[22] Guo M, Wang R, Wang J, Hua K, Wang Y, Liu X, Yao S. ALT1, a Snf2 family chromatin remodeling ATPase, negatively regulates alkaline tolerance through enhanced defense against oxidative stress in rice. PLoS One 2014;9:e112515.
[23] Liu H, Li J, Wang S, Hua J, Zou B. CHROMATIN REMODELING 11-dependent nucleosome occupancy affects disease resistance in rice. Plant Physiol 2023;193:1635–51.
[24] Flicek P, Amode MR, Barrell D, Beal K, Brent S, Chen Y, Clapham P, Coates G, Fairley S, Fitzgerald S, et al. Ensembl 2011. Nucleic Acids Res 2011;39:D800–6.
[25] Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N, et al. Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res 2012;40:D1178–86.
[26] Jin J, Tian F, Yang DC, Meng YQ, Kong L, Luo J, Gao G. PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. Nucleic Acids Res 2017;45:D1040–5.
[27] Li Z, Zhao Y, Liu X, Peng J, Guo H, Luo J. LSD 2.0: an update of the leaf senescence database. Nucleic Acids Res 2014;42:D1200–5.
[28] Yang J, Liu Y, Yan H, Tian T, You Q, Zhang L, Xu W, Su Z. PlantEAR: functional analysis platform for plant EAR motif-containing proteins. Front Genet 2018;9:590.
[29] UniProt C. UniProt: the universal protein knowledgebase in 2021. Nucleic Acids Res 2021;49:D480–9.
[30] Marakulina D, Vorontsov IE, Kulakovskiy IV, Lennartsson A, Drablos F, Medvedeva YA. EpiFactors 2022: expansion and enhancement of a curated database of human epigenetic factors and complexes. Nucleic Acids Res 2023;51:D564–70.
[31] Liu L, Zhen XT, Denton E, Marsden BD, Schapira M. ChromoHub: a data hub for navigators of chromatin-mediated signalling. Bioinformatics 2012;28:2205–6.

[32] Ryan DP, Owen-Hughes T. Snf2-family proteins: chromatin remodellers for any occasion. Curr Opin Chem Biol 2011;15:649–56.
[33] Hu Y, Zhu N, Wang X, Yi Q, Zhu D, Lai Y, Zhao Y. Analysis of rice Snf2 family proteins and their potential roles in epigenetic regulation. Plant Physiol Biochem 2013;70:33–42.
[34] Zhang D, Gao S, Yang P, Yang J, Yang S, Wu K. Identification and expression analysis of Snf2 family proteins in tomato (solanum lycopersicum). Int J Genom 2019;2019:5080935.
[35] Shang JY, He XJ. Chromatin-remodeling complexes: conserved and plant-specific subunits in Arabidopsis. J Integr Plant Biol 2022;64:499–515.
[36] Oughtred R, Stark C, Breitkreutz BJ, Rust J, Boucher L, Chang C, Kolas N, O'Donnell L, Leung G, McAdam R, et al. The BioGRID interaction database: 2019 update. Nucleic Acids Res 2019;47:D529–41.
[37] Shamimuzzaman M, Gardiner JM, Walsh AT, Triant DA, Le Tourneau JJ, Tayal A, Unni DR, Nguyen HN, Portwood 2nd JL, Cannon EKS, et al. MaizeMine: a data mining warehouse for the maize genetics and genomics database. Front Plant Sci 2020;11:592730.
[38] Papatheodorou I, Moreno P, Manning J, Fuentes AM, George N, Fexova S, Fonseca NA, Fullgrabe A, Green M, Huang N, et al. Expression Atlas update: from tissues to single cells. Nucleic Acids Res 2020;48:D77–83.
[39] Edgar RC. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinforma 2004;5:113.
[40] Potter SC, Luciani A, Eddy SR, Park Y, Lopez R, Finn RD. HMMER web server: 2018 update. Nucleic Acids Res 2018;46:W200–4.
[41] Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, et al. InterProScan 5: genome-scale protein function classification. Bioinformatics 2014;30:1236–40.
[42] Retief JD. Phylogenetic analysis using PHYLIP. Methods Mol Biol 2000;132:243–58.
[43] Persson E, Sonnhammer ELL. InParanoid-DIAMOND: faster orthology analysis with the InParanoid algorithm. Bioinformatics 2022;38:2918–9.
[44] Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol 2019;20:238.
[45] Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, Xia R. TBtools: an integrative toolkit developed for interactive analyses of big biological data. Mol Plant 2020;13:1194–202.
[46] Chen G, Mishina K, Zhu H, Kikuchi S, Sassa H, Oono Y, Komatsuda T. Genome-wide analysis of Snf2 gene family reveals potential role in regulation of spike development in barley. Int J Mol Sci 2022;24.
[47] Song ZT, Liu JX, Han JJ. Chromatin remodeling factors regulate environmental stress responses in plants. J Integr Plant Biol 2021;63:438–50.
[48] Hernandez-Garcia J, Diego-Martin B, Kuo PH, Jami-Alahmadi Y, Vashisht AA, Wohlschlegel J, Jacobsen SE, Blazquez MA, Gallego-Bartolome J. Comprehensive identification of SWI/SNF complex subunits underpins deep eukaryotic ancestry and reveals new plant components. Commun Biol 2022;5:549.